E-Lamp Prototype System

CMU E-Lamp Team

Carnegie Mellon University

Introduction

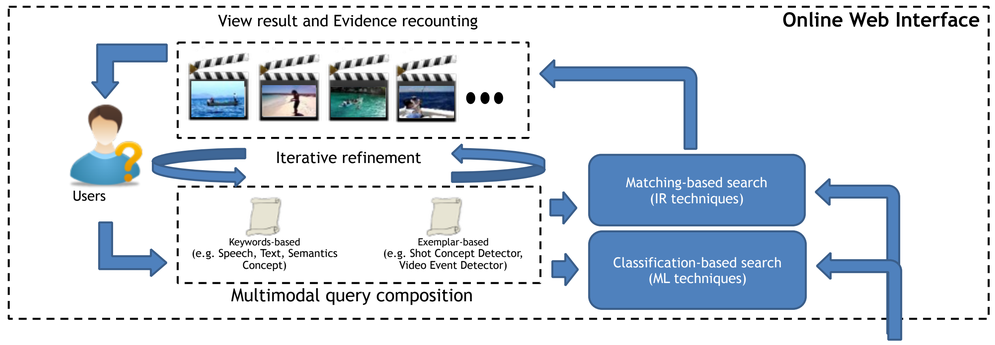
The E-Lamp Prototype System helps users to search for video events or create content tags by using the CMU MED14 system. The system can be divided into two parts:

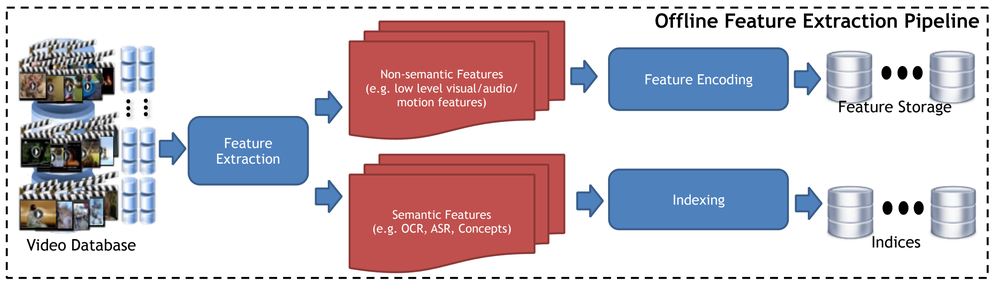
1. Online Web Interface: The interface consists of three modules: **collection tag summary**, **multimodal query generation** and **search result visualization**.
2. Offline Feature Extraction Pipeline: The prototype builds on semantic concepts as metadata and 19 non-semantic low-level features.

Collection Tag Summary

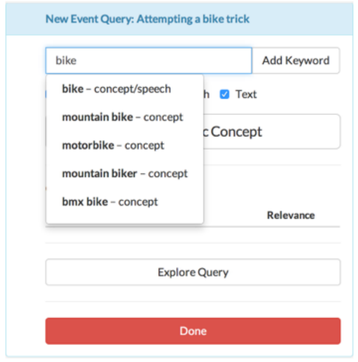
The system summarizes high frequency speech/text tags and high confidence visual content tags in each collection with word clouds. The word size is proportional to the tag frequency/ confidence, clicking on the tag will show the top 100 videos.

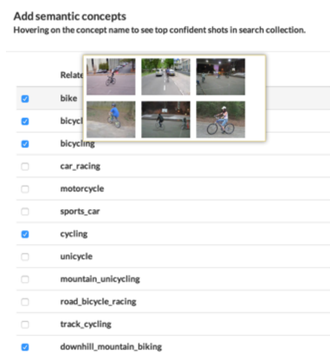
System Architecture

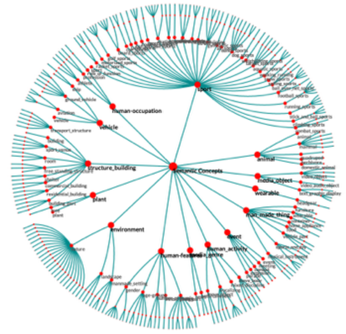




Multimodal Query Generation

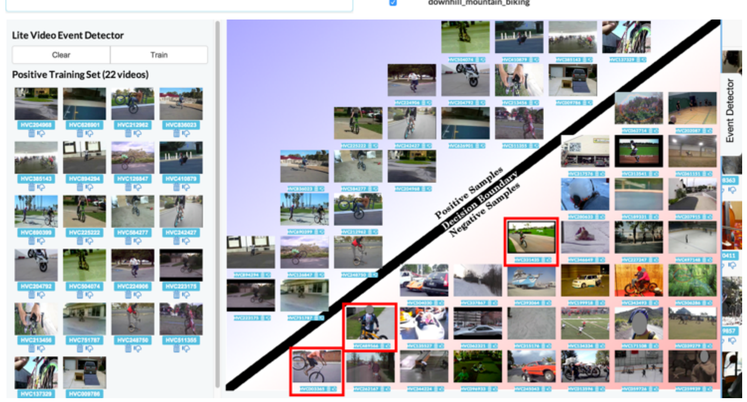


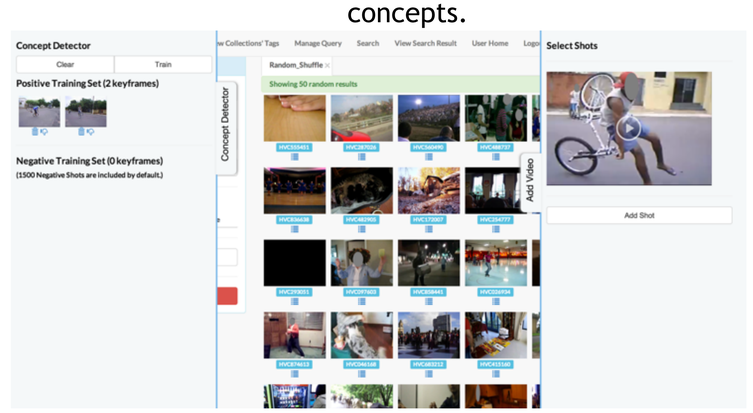




How to create a good query?

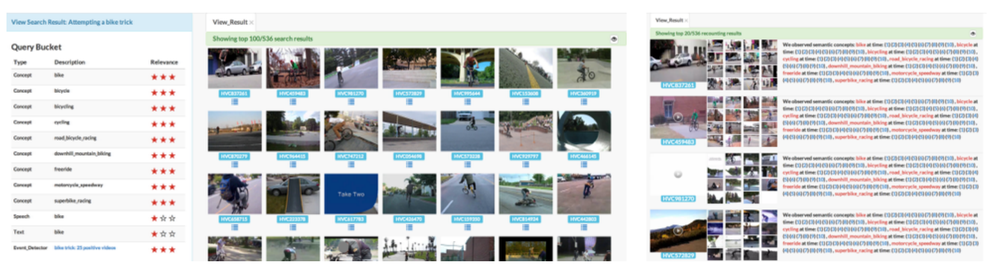
1. Start with keyword-based query
1.1. Input speech, text or concept keywords.
1.2. Select related concepts. Select reliable concepts by looking at the high confidence shot examples.
1.3. Browse the concept hyper-tree to find additional concepts.





2. Create exemplar-based query. For event query generation, users can build a detector for complete event videos. Alternatively, in content tag generation, users can choose individual shots to build a shot level concept detectors.
2.1. Users can then interactively refine the detector by utilizing model visualization and exploring the query. For example, in model visualization, a user could add false negatives (in red box) back into the positive training set.

Search Result Visualization



Compact mode shows example thumbnails and image of top video search results.

Detail mode shows speech, text or semantic evidence for recounting of top video search results.

Conclusion

- We built a prototype system that enables users to search video events in real time. The prototype system can achieve 86% performance(MAP) compared to our full MED system on MEDTEST 14(E021-E030). The average system compute time for training and search is around three minutes.
- Over 2500 semantic concepts have been incorporated for search. Users can also build a concept detector online and use it as a new content tag.
- We developed interfaces to help users build and refine a multimodal query, and achieve reasonable search results on new video events.

Acknowledgement

This work has been supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center contract number D11PC20068. The U.S. government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.